

Unbounded memory advantage in stochastic simulation using quantum mechanics.

Andrew J. P. Garner,^{1,2,*} Qing Liu,³ Jayne Thompson,¹ Vlatko Vedral,^{4,1,5,2} and Mile Gu^{3,6,1,†}

¹Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore

²Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084, China

³School of Physical and Mathematical Sciences, Nanyang Technological University, 639673, Singapore

⁴Atomic and Laser Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, United Kingdom.

⁵Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542

⁶Complexity Institute, Nanyang Technological University, 639673, Singapore.

(Dated: 16th September 2016)

Simulating the stochastic evolution of real quantities on a digital computer requires a trade-off between the precision to which these quantities are approximated, and the memory required to store them. The statistical accuracy of the simulation is thus generally limited by the internal memory available to the simulator. Here, using tools from computational mechanics, we show that quantum processors with a fixed finite memory can simulate stochastic processes of real variables to arbitrarily high precision. This demonstrates a provable, unbounded memory advantage that a quantum simulator can exhibit over its best possible classical counterpart.

PACS numbers: 03.67.-a, 02.50.Ey, 05.20.-y

Many macroscopic processes we wish to simulate involve the dynamics of real numbers. The dynamical properties we wish to track (e.g. the position of an object) can take on almost any number, seemingly without noticeable quantization until one goes down to the Planck scale. The simulation of such processes necessitates compromise between the resources allocated and the precision with which we track such properties. Clever implementations to this problem, such as the floating point format [1], form the heart of modern computing technology – but all subscribe to the same trade-off: treating a quantity with higher precision requires the allocation of more memory. To perfectly replicate the future statistics of a continuous variable dynamical system exactly would inevitably require unbounded memory.

The advent of quantum technology, however, opens new possibilities. Not only has this technology shown great potential in solving problems many consider classically intractable [2–6], it has demonstrated the capability to greatly reduce the amount of information one needs to send in certain tasks requiring communication between distributed parties [7–9]. Could the memory required by a quantum machine that simulates dynamical processes likewise scale much more favourably with precision?

Here, we consider the simulation of a class of stochastic systems involving the dynamics of parameters that takes on real numbers. Classical simulation of such processes digitally involves ‘coarse-graining’: the parameter at each point in time is approximated to n bits of precision at some memory cost that scales linearly with n . We construct quantum simulators that exhibit *unbounded* advantage. The quantum simulator can exactly replicate the statistics of a n bit classical simulator for arbitrarily large n using a bounded amount of memory. Thus,

quantum simulators can side-step the precision-memory tradeoff – finite quantum memory can simulate such processes to arbitrary fixed precision.

This unbounded divergence has practical and foundational consequences. Practically, it suggests that quantum processors may be increasingly advantageous as we wish to simulate ever more memory-intensive systems, such as those arising from big data sets. Foundationally, since the minimal memory required to simulate a process is a well established measure of its complexity [10–20], our work suggests that there are certain processes which grow unboundedly in complexity, but yet remain simple to an observer with quantum capabilities.

Cyclic random walks. Consider a small bead located on a circular ring (as per figure 1). Its position can always be described by some real number $y \in [0, 1)$. At each discrete time $t \in \mathbb{Z}$, the bead’s position is stochastically perturbed. This perturbation is described by a real random variable X that is governed by a continuous probability density function $P(X)$; such that

$$Y^{t+1} = \text{frac}[Y^t + X], \quad (1)$$

where Y^t represents the random variable that governs the location of the bead at time t , and $\text{frac}[y] = y - [y] \in [0, 1)$.

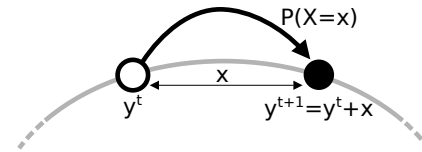


Figure 1: **Cyclic random walk.** At each time step, the system stochastically hops from state $y^t \in [0, 1)$ to $y^{t+1} = \text{frac}[y^t + x]$. As x is chosen according to the real random variable X , the current value of the system is itself described by a sequence of real random variables $\{Y^t\}_{t \in \mathbb{Z}}$ that satisfy $Y^{t+1} = \text{frac}[Y^t + X]$.

* ajpgarner@nus.edu.sg

† gumile@ntu.edu.sg

$[0, 1)$ denotes the fractional part of y ($\lfloor y \rfloor$ is the lowest integer below or equal to y); and captures the fact that y and $y + m$ represent the same location on the ring for any integer m . We refer to $P(X)$ as the *shift function*, and assume the process is stationary in the sense that $P(X)$ has no explicit dependence on t . Although we motivate this process as a bead on a ring for clarity, the same formalism describes a diverse range of systems undergoing *cyclic random walks*, such as the azimuthal motion of gas molecules diffusing in an annular tube, or the position of a single electron travelling through an electric circuit with constant resistance.

We capture the dynamics of Y formally using the framework for describing stochastic processes. In general, a stochastic process \mathcal{P} is characterized by a bi-infinite sequence of random variables $\{Y^t\}_t$, that governs its value at each discrete time $t \in \mathbb{Z}$. For convenience, we often segregate past and future values, such that $\vec{Y} = \dots Y^{-1}Y^0$ and $\vec{Y} = Y^1Y^2\dots$ respectively govern the values in the past and future with respect to time $t = 0$. The cyclic random walk above is then entirely captured by the joint probability distribution $P(\vec{Y}, \vec{Y})$ such that for any instance of the process with past values \vec{y} , future values \vec{y} will be observed with probability $P(\vec{Y} = \vec{y} | \vec{Y} = \vec{y})$.

Here, we consider the simulations of the above process to ever increasing precision. We adopt a natural technique of discretizing a continuous process, by introducing a family of stochastic processes $\{\mathcal{P}_n\}$ that describe discrete approximations of this process, where in each the position of bead is represented to n bits of precision by a n -digit binary number. This is done by limiting y to a discrete set of $N = 2^n$ equally-spaced values, $y_j = j/2^n$ (for $j = 0$ to $N - 1$). At each time-step, the probability that a bead in discrete location y_j transitions to y_k , is given by the probability p_{jk} that a bead initially at y_j will transition to any value of y whose n bit binary representation is y_k . That is

$$p_{kj} = P(Y^{t+1} = y \in \mathcal{I}_k | Y^t = y_j) \quad (2)$$

where $\mathcal{I}_k = \{y : |y - y_k| < \frac{1}{2^n}\}$ represents the interval on the ring that is ‘rounded to’ y_k . This results in a Markovian stochastic process that emits a symbol from the finite alphabet $\{y_k\}$ at each time-step, whose dynamics are governed by the stochastic matrix with elements p_{jk} . As $n \rightarrow \infty$, the statistics of \mathcal{P}_n approach that of \mathcal{P} ; at the potential cost of tracking more information¹.

Classical simulation costs scale with precision.

We can formally describe simulators using the tools of computational mechanics [10–13]. A simulator of a process is a device whose future output behaviour conditioned on any particular past should be statistically indistinguishable to the process itself. Specifically, let the state of the simulator at each time-step be s^t , such that

at the subsequent time-step it can output y^{t+1} and transition to state s^{t+1} . For this device to be a statistically faithful simulation of a process $P(\vec{Y}, \vec{Y})$, we require that:

1. For each specific past \vec{y} at each time t , we can deterministically configure the device using a function f into some state $s = f(\vec{y})$, such that it will produce future outputs \vec{y} with probability $P(\vec{Y} = \vec{y} | \vec{Y} = \vec{y})$.
2. If a simulator is in state $s^t = f(\vec{y})$ at time t , and outputs y^t in the subsequent time-step, its internal state must then transition to $s^{t+1} = f(\vec{y}y^t)$.

The first condition ensures the simulator can be initialized to simulate desired conditional future statistics; the second that a correctly initialized simulator continues to exhibit statistically correct statistics at every time-step. The entropy of the simulator’s internal state then represents the memory cost of the simulation, and is clearly bounded below by the information entropy of the random variable $S = f(Y)$ that governs it. Thus, physically a simulator can be viewed as communication channel in time: it represents the exact object Alice must give to Bob at each time-step that captures sufficient past information for Bob to replicate the processes conditional future behaviour. f is known as the *encoding function*, that describes how the past is encoded within the channel.

The memory cost of the provably-optimal classical simulator – known as the *statistical complexity*, C_μ – is extensively studied in complexity science [10]. This value captures the absolute minimum memory any classical simulator of a process must store, and thus is a prominent quantifier of a process’s structure and complexity (e.g. [13–20]). Such an optimal simulator can be explicitly constructed, and corresponds to the simulator that stores in its internal memory the *causal states* of the process [10, 11]: defined by an encoding function f such that $f(\vec{y}) = f(\vec{y}')$ if and only if $P(\vec{Y} | \vec{Y} = \vec{y}) = P(\vec{Y} | \vec{Y} = \vec{y}')$ (i.e. the conditional futures of \vec{y} and \vec{y}' coincide).

In our cyclic random walks, each \mathcal{P}_n is Markovian: the statistics of future outcomes depend only on the most-recent value of Y_t . When discretized, the causal states are thus typically in one-to-one correspondence with the 2^n discrete values that Y can take². That is, \mathcal{P}_n has 2^n causal states, labelled $\{s_j\}_{j=0}^{2^n-1}$, where s_j corresponds to the set of pasts ending in $Y^0 = y_j$. At the steady state, $P(S = s_i) = \frac{1}{N}$, as all causal states occur with equiprobability, and thus classical statistical complexity

$$C_\mu = n, \quad (3)$$

scales linearly with the precision.

¹ An alternative discretization is to calculate the the transition probabilities by assuming the initial value of y^t is uniformly distributed in \mathcal{I}_j . This yields asymptotically identical statistics as $N \rightarrow \infty$, and does not change the results of this article.

² There are exceptions, such as when $P(x) = 1$ for $x \in [0, 1)$, and the system jumps to a completely random point at each time-step; here there is only one causal state for all N , because the current position no longer affects the future outcomes at all.

Quantum simulations are memory-efficient. Quantum processors have recently been demonstrated the capability to simulate stochastic processes with less memory than any classically possible [21–25]. Here, we construct an explicit quantum simulator for the cyclic random walk. Instead of storing each causal states s_i directly, our quantum simulator stores a corresponding quantum state

$$|S_j\rangle = \sum_{k=0}^{N-1} \sqrt{p_{kj}} |k\rangle, \quad (4)$$

where $\{|k\rangle\}$ form an orthonormal basis.

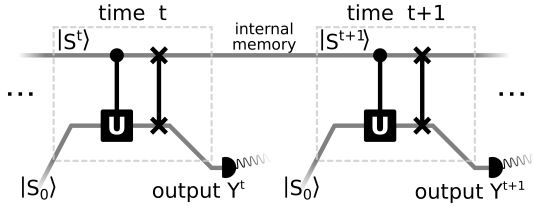


Figure 2: Circuit for memory-efficient quantum simulation. The above circuit illustrates a quantum simulator that samples $P(\tilde{Y}|\tilde{y})$ when supplied with the appropriate quantum state $|S^t\rangle$ that encodes the past. At $t = 0$, an ancilla system, initialized in state $|S_0\rangle$, is fed into the simulator. A controlled unitary is then enacted such that $U : |j\rangle |S_0\rangle \rightarrow |j\rangle |S_j\rangle$ for each j . The state of the ancillary system and memory are then coherently swapped, and the ancilla is then emitted as output. Measurement of the ancilla then correct samples \tilde{Y}^1 . Iteration of this procedure then generates output behaviour statistical identical to that of the original process.

The stationary state of the quantum simulator is then given by the *quantum ensemble state* $\rho = \frac{1}{N} \sum_j |S_j\rangle \langle S_j|$ (as all quantum states occur with equiprobability). Thus the memory required to store these states is given by the von Neumann entropy given $H_Q := -\text{Tr}(\rho \log \rho) = -\sum_k \lambda_k \log \lambda_k$, where λ_k are the eigenvalues of ρ . The key improvement here is that $\{|S_j\rangle\}$ are not in general mutually orthogonal, and thus H_Q is generally less than C_μ . Nevertheless a quantum circuit (outlined in figure 2 – with details in the *Technical Appendix*) acting on these quantum states will produce statistically identical outputs to the classical simulator.

Unbounded advantage of quantum memory.

We now come to the main claim of our paper: there are stochastic processes that can be simulated to infinite precision using a finite amount of quantum memory.

Explicitly, we show that for certain cyclic processes, the quantum ensemble state’s eigenvalues $\{\lambda_k\}_{k=0\dots N-1}$ to satisfy $\lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} -\lambda_k \log \lambda_k = \Omega$ for some finite value Ω . Our result relies on first observing that the eigenvalues λ_k can be directly related to transition probabilities $\{p_{jk}\}$ via the relation

$$\lambda_k = \frac{1}{N} \mathcal{F}[\sqrt{p_{j0}}] \mathcal{F}[\sqrt{p_{(N-j)0}}], \quad (5)$$

where \mathcal{F} denotes the *discrete Fourier transform*, $\mathcal{F}(x_j) = \sum_{j=0}^{N-1} x_j \exp\left(\frac{-2\pi i}{N} jk\right)$. (The proof relies on invoking the cyclic symmetry of the process – and hence of the transition probabilities – and is explicitly derived in *Technical Appendix*). The spread p_{j0} (as a function of j) is an indicator of how quickly a particle diffuses in the random walk. Thus, the Fourier-like relation between p_{j0} and λ_k indicates an inverse relationship between the amount of diffusion in the cyclic process and the spread of eigenvalues. The greater the variance of X , the more quickly a particle diffuses, and the smaller the spread of λ_k – resulting in a reduced quantum memory requirement. We now show that for some natural examples, this reduction is sufficiently large that H_Q remains bounded for all n (as illustrated in figure 3).

Example 1: Gaussian noise. A cyclic process rotating at a constant rate subject to Gaussian noise has a shift function given by a Gaussian distribution $G_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ about mean μ with standard deviation σ . Here, μ characterises the average velocity (in terms of the variable’s mean displacement per time-step), and σ the size of the fluctuations. When $\mu = 0$, this process corresponds to Gaussian diffusion. For our analysis, we take $\sigma \ll 1$ and thus ignore fluctuations where the particle travels more than a complete loop around the ring in a single time-step (a value of $\sigma = 0.1$ ensures that such events are less likely than one part in a million.)

As can be seen in figures 3a and 3c, as the desired precision increases, the memory cost of simulating this process quickly converges onto a constant determined by the fluctuation strength σ ; ultimately, *infinite-precision simulation is possible using only a finite quantum memory*. This behaviour may be understood analytically by seeing that for large N , the eigenvalues associated with the quantum simulator’s internal memory are also given by samples from a Gaussian distribution: $\lambda_k = G_{0, \frac{1}{4\pi\sigma}}(k)$ for $k = -\frac{N}{2}, \dots, \frac{N}{2} - 1$, where for convenience we have cyclicly offset the label of the eigenvalues’ indices by N (proof in *Technical Appendix*). This demonstrates that increasing σ tightens the spread of eigenvalues, and thus reduces the memory requirement for the quantum simulator.

In the *Technical Appendix*, we prove that as the precision $n = \log N$ increases, the sum $\lim_{N \rightarrow \infty} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \lambda_k \log \lambda_k$ converges on a finite value, bounded (in bits) by

$$H_Q \leq \frac{1}{2 \ln 2} - \left(1 + 4\sqrt{2\pi}\sigma\right) \log_2 2\sqrt{2\pi}\sigma \quad (6)$$

Thus, for any fixed $0 < \sigma \ll 1$, the Gaussian random walk may be simulated to arbitrarily high precision using a quantum simulator of bounded entropy. Moreover, this also implies an unbounded divergence between the classical and the *quantum statistical complexity* [22, 26] C_Q , which is upper bounded by H_Q .

Example 2: Uniform white noise. In the second example, we consider a particle that is perturbed by uni-

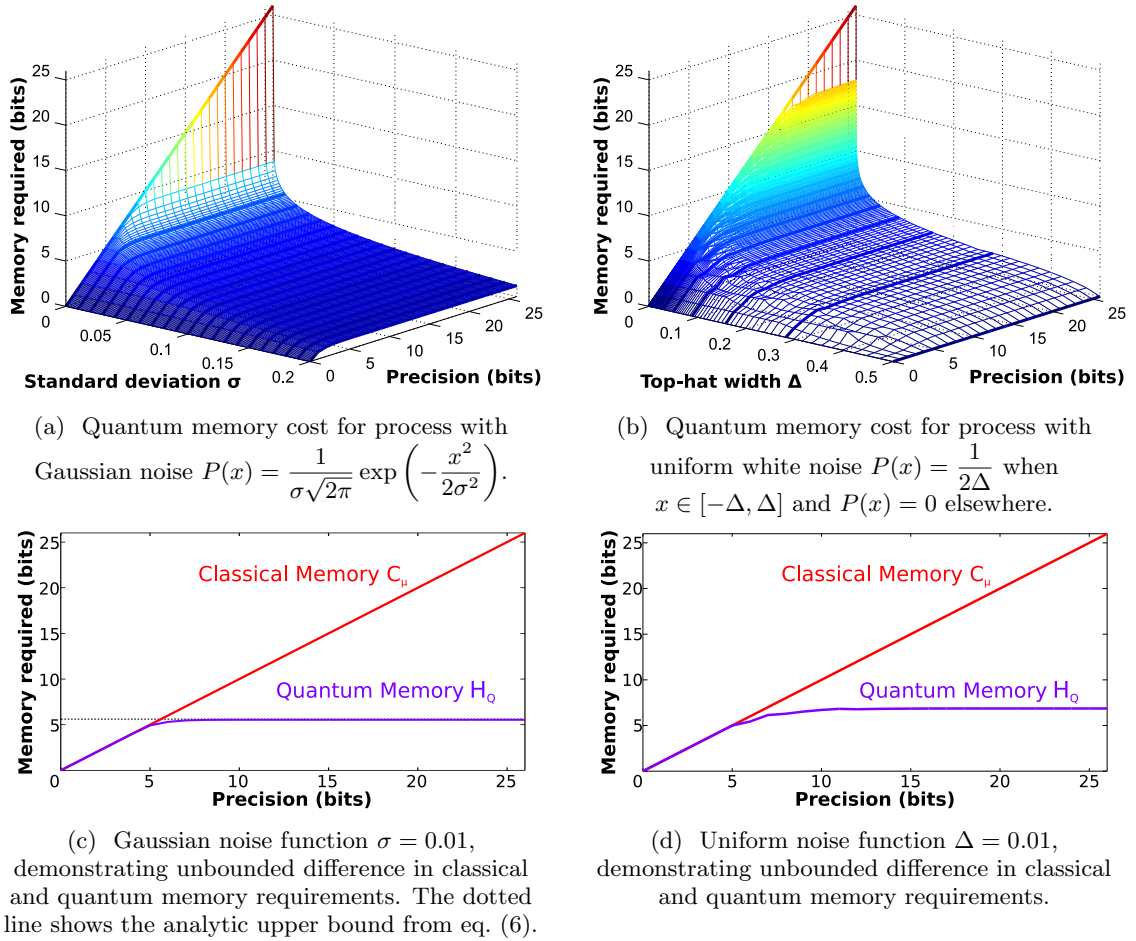


Figure 3: **Bounded quantum memory costs for unbounded precision.** The memory required to simulate a cyclic random walk is plotted against the precision N for the *Gaussian* and *top-hat* shift functions. In both examples, the quantum simulator has an unbounded memory advantage – the classical cost scales as $\log N$ whilst the quantum cost converges upon a constant value. The more rapidly the shift function diffuses X , the lower the limiting quantum memory requirement.

formly distributed noise. At each time-step, the particle can move anywhere in the range of $\mu \pm \Delta$ from its current position with uniform probability, where $\Delta < \frac{1}{2}$. Again, μ characterises the average velocity, and here Δ the size of the fluctuations. The associated shift function is a *top-hat function*, that has a uniform value of $\frac{1}{2\Delta}$ in the range $x \in [\mu - \Delta, \mu + \Delta]$ and 0 everywhere else.

The entropy of the quantum simulator, H_q is plotted for various precision in figures 3b and 3d. We see that for any fixed $\Delta > 0$, the quantum memory required by our simulator converges to a bounded value. As in the Gaussian scenario, the quantum simulator can replicate a classical simulation to any given precision using with finite entropy. In the *Technical Appendix*, we prove this analytically. We show that as $N \rightarrow \infty$, the entropy remains finite, and is bounded above by $H_Q \leq \frac{1.894}{\sqrt{\Delta}} + 3.067$. In particular, for large N , the eigenvalues of the relevant ensemble state obey $\lambda_k = 2\Delta \text{sinc}^2(2k\Delta)$ for $k = -\frac{N}{2}, \dots, \frac{N}{2} - 1$, where $\text{sinc}(x)$ is the *normalized sinc function*, $\text{sinc}(x) := \frac{1}{\pi x} \sin(\pi x)$. Larger values Δ will result in a smaller spread of eigenvalues, and result in smaller H_q . For any given $\Delta > 0$ the entropy is finite in the limit $N \rightarrow \infty$. This establishes a second natural

example where the quantum simulator can demonstrate an unbounded memory advantage over its best possible classical counterpart.

The origin of quantum advantage. The source of classical inefficiency can be understood by considering dynamics on causal states. Consider two instances of \mathcal{P}_n . One where $Y^0 = y_j$, and the other where $Y^0 = y_{j+1}$. As their conditional futures differ, a classical simulator must be configured differently for each instance (corresponding to being initialized in one of two different causal states, s_j or s_{j+1}). Nevertheless, there is finite probability that at the next time-step, both instances of the process emit the same output (up to precision n). Should this happen, we would not be able to retrodict which causal state we came from. That is, there is some probability that the distinction between s_j and s_{j+1} will never be reflected in the future statistics of the process – a phenomenon known as *crypticity* [23, 27]). As n increases, this occurs with greater likelihood (tending to unit probability as $n \rightarrow \infty$), and thus proportionally more information is wasted.

Quantum simulators compensate for this waste by mapping these causal states to non-orthogonal quantum states. The quantum state (eq. (4)) associated with neighbouring causal states ($|S_j\rangle$ and $|S_{j+1}\rangle$) also become increasingly similar with increasing n – resulting in progressively greater savings. Consider the Gaussian scenerio, where H_q is bounded by equation (6). For small σ , the memory cost scales as $-\log_2 \sigma$, such that halving the variance of fluctuation's at each time-step adds one bit to the memory cost of the quantum simulator. The standard deviation of the shift function has set an effective length scale over which the system must be simulated classically. The statistical behaviour of future outputs from two systems that are initially prepared in points separated by more than one standard deviation are typically distinguishable, and so these points must be stored as nearly-orthogonal quantum states at some memory cost. On the other hand, when two points are initially closer than the standard deviation scale, the probability that they could be distinguished by their future behaviour diminishes, and they may be represented by increasingly overlapping quantum states. In this regime, a fixed finite memory can accommodate any desired precision.

Outlook and discussion. In this article, we presented a task in which quantum mechanics has an unbounded memory advantage over the most memory-efficient classical alternative: the simulation of a classical cyclic stochastic process. We found that the classical simulator has a memory requirement that scales linearly with the precision required, while the quantum simulator's requirement may be bounded by a finite value, even at arbitrarily-high fixed precision. This establishes a rare scenario where the scaling advantage of quantum processing can be provably established.

This finding leads a number of natural open questions – the first being of generality. Certainly, the examples presented are sufficiently simple that such divergences are unlikely to be merely a mathematical oddity. The unbounded quantum advantage relies on $\{\mathcal{P}_n\}$ having two

properties: (a) the number of causal states grows with n , and (b) the conditional futures between different causal states converges sufficient quickly with n . If these conditions can be formalized, we may be able to establish similar divergences much more general scenarios, such as the simulation of non-Markovian or non-cyclic processes. Meanwhile the inefficiency of classical simulators have show to directly results in unavoidable increased heat dissipation [28–30]. This hints that quantum processing may allow significant energetic savings for stochastic simulation, especially for systems that scale in complexity.

On a more foundational level, the statistical complexity is often regarded as a fundamental measure of a processes's intrinsic structure – the rationale being that it quantifies the minimal amount of causal information one must postulate to understand the process's future behaviour. The measure has been applied to understand structure within diverse settings, from the dynamics of neurons [14] to quantifying self-organization [15], among other examples [16–20]. The discovery of more efficient quantum models led to great interest in the idea that complexity depends on what sort of information we use to observe a system [22, 26]. In this context, our results establish a family of processes can look ever more complex classically, but remain simple quantum-mechanically. It would be fascinating to see if such divergences can be found in existing studies, such that systems currently considered to be immensely complex may look rather simple through the lens of quantum theory.

Acknowledgements. We thank Thomas Elliott, David Garner, Jan-Åke Larsson, and Chengran Yang for helpful comments and discussions. We gratefully acknowledge funding from the John Templeton Foundation Grant 53914 “*Occam’s Quantum Mechanical Razor: Can Quantum theory admit the Simplest Understanding of Reality?*”; the Foundational Questions Institute; the Ministry of Education in Singapore, the Academic Research Fund Tier 3 MOE2012-T3-1-009; and the the National Research Foundation of Singapore (Award Nos. NRF-NRFF2016-02 and NRF-CRP14-2014-02).

TECHNICAL APPENDIX

Classical costs from computational mechanics.

We here present some minimal details from the mathematical framework of computational mechanics [10–13] to substantiate the claim that the classical simulator’s minimal memory cost is equal to the precision $\log N$.

In computational mechanics, the evolution of a dynamical property (over domain \mathcal{Y}) is characterised by a discrete-time stochastic process \mathcal{P} , written as bi-infinite sequence of random variables $\{Y^t\}_{t \in \mathbb{Z}}$, where each random variable Y^t governs the value $y^t \in \mathcal{Y}$ of the dynamical property at time t . The statistical behaviour of a process may be represented in a *causal* manner by writing it as the conditional probability distribution $P(\vec{Y}_t | \vec{Y}_t)$, where $\vec{Y}^t = Y^{t+1}Y^{t+2} \dots$ is the infinite string of random variables occurring after time t , and $\vec{Y}^t = \dots Y^{t-1}Y^t$ is the infinite string of random variables occurring before (and including) time t . For *stationary* processes (such as the time-independent cyclic random walks described in this article), this distribution has no explicit time dependence, so we omit the superscript t .

A faithful *simulation* of process \mathcal{P} is a machine (or program) that, having been initialized in accordance with the observation of past \vec{y}^t , then generates a series of outputs \vec{y}_t according to the distribution $P(\vec{Y}^t = \vec{y}^t | \vec{Y}^t = \vec{y}^t)$. Since, storing infinite string \vec{y}^t may require an unbounded amount of memory, one instead configures the internal state of the simulator s (over configuration space \mathcal{S}) according to some function $s = f(\vec{y})$, satisfying $P(\vec{Y}^t = \vec{y}^t | S = s) = P(\vec{Y}^t = \vec{y}^t | \vec{Y}^t = \vec{y}^t)$, where $S = f(\vec{Y})$ is the random variable describing the internal state of the simulator (formed by applying the function f on each variate of \vec{Y}). Moreover, once initiated into state s^t , when the simulator outputs y^t in the subsequent time-step, its internal state must then transition to the state $s^{t+1} = f(\vec{y}y^t)$ (where $\vec{y}y^t$ indicates the concatenation of y^t to the end of string \vec{y}).

The the memory cost of such a simulator is given by the information entropy of S , $H(S) = -\sum_{s_i \in \mathcal{S}} P(S = s_i) \log P(S = s_i)$. The function f that minimizes this classically corresponds to identifying the *causal state* of a particular past [10, 11], defined by the equivalence relationship: $\vec{y} \sim_\epsilon \vec{y}'$ for pasts \vec{y} and \vec{y}' if and only if $P(\vec{Y} = \vec{y} | \vec{X}_t = \vec{y}) = P(\vec{Y} = \vec{y}' | \vec{X}_t = \vec{y}')$ for all possible future values $\vec{y} \in \vec{Y}$. The causal states are unique for any given process, and so their entropy $H(S)$ is a property of the process itself known as its *statistical complexity* C_μ , capturing the intuition that a more complex process requires more memory to simulate.

For Markovian processes, such as discussed in this article, the number of causal states required is equal to the number of unique rows in the stochastic matrix describing the evolution. When these rows are generated by the discretization of a continuous process into N divisions – such as when they are derived from the cyclic walk’s shift function $P(X)$ – the number of states will be

equal to N , except for very specific (e.g. pathologically fractal) choices of $P(X)$ and N . Since by symmetry the probability of the simulator being in any particular state is equal, the classical memory cost of a simulator hence scales with the number of sites as $\log N$, or linearly with the precision $n = \log_2 N$.

Details of the quantum circuit in figure 2. Let us consider figure 2 in more depth (see also [21]). The circuit consists of one persistent internal memory state, and as an “output tape” a line of quantum states, which are fed into the system one at a time. Suppose each state on the output tape is initialized into some arbitrary state $|\phi\rangle$. For any two quantum states $|x\rangle$ and $|y\rangle$ in the same Hilbert space, it is always possible to construct a unitary transformation V such that $V|x\rangle = |y\rangle$. This will be of the form $|y\rangle\langle x| + \sum_i |y'_i\rangle\langle x'_i|$ where $|x'_i\rangle$ are states orthogonal to each other and to $|x\rangle$, and $|y'_i\rangle$ are states orthogonal to each other and to $|y\rangle$. Thus, in the joint Hilbert space $\mathcal{H}^N \otimes \mathcal{H}^N$ of two quantum systems of dimension N , it is possible to build a “controlled” unitary operation U containing the elements $|j\rangle\langle j| \otimes |\psi_j\rangle\langle \phi|$ for every $|\psi_j\rangle$ in an arbitrary (generally non-orthogonal) set of states $\{|\psi_j\rangle\}_{j=0 \dots (N-1)}$. [Note: the orthogonality of $\{|j\rangle\}$ allows us to pairwise use the above construction for each $|\psi_j\rangle$.]

For a Markovian process discretized such that the stochastic matrix with elements p_{jk} describes its evolution, the above prescription supplies the unitary operation required for our quantum simulator when we set each $|\psi_j\rangle = |S_j\rangle = \sum_{k=0}^{N-1} \sqrt{p_{kj}} |k\rangle$, as per eq. (4) (states $\{|k\rangle\}$ and $\{|j\rangle\}$ are in the same basis).

We may now evaluate the action of a single time-step (grey dashed box within figure 2). Here, the joint Hilbert space corresponds to that of the internal memory together with the output tape. In the figure, we explicitly wrote the initial state of the output tape as $|\phi\rangle = |S_0\rangle$, but this is arbitrary; any $|\phi\rangle$ could be made into $|S_0\rangle$ by acting on it first with a unitary gate containing $|S_0\rangle\langle \phi|$. At the start of a time step, the internal memory is in state $|S^t\rangle = |S_j\rangle = \sum_{k=0}^{N-1} \sqrt{p_{kj}} |k\rangle$. Hence, the joint state of the memory and output tape is initially $|S_j\rangle \otimes |\phi\rangle$. After the controlled unitary is applied will be in the entangled state $\sum_k \sqrt{p_{kj}} |k\rangle \otimes |S_k\rangle$. Applying a coherent swap operation (i.e. exchanging the labels of the Hilbert spaces) will take this joint state to $\sum_k \sqrt{p_{kj}} |S_k\rangle \otimes |k\rangle$ – the state of the system at the end of the grey box.

The tape system is then ejected from the simulator. If one were to measure this state in the $\{|k\rangle\}$ basis, one projects onto state $|k\rangle$ with probability p_{kj} , and hence the output statistics of this measurement match that of the process being simulated. Moreover, after measuring, due to the entanglement, we know that when $|k\rangle$ is measured, the internal memory must be in state $|S_k\rangle$, which is exactly the quantum state that would have been prepared if we had mapping the output statistics onto a classical causal state and then prepared $|S_k\rangle$ directly. Hence, the quantum circuit in figure 2 can function as a discretized simulator for a Markovian process.

However, it is very important to note that there is no need whatsoever to measure the output tape $|k\rangle$ for the quantum simulator to continue functioning. If it suits one's purpose to store the output states in quantum memory (e.g. to perform further quantum information processing on the output data), then the quantum simulator still functions correctly. In this mode of operation, the measurements can be omitted from figure 2, and after M steps, the simulator would have produced the entangled state

$$|\Phi\rangle = \sum_{i_1} \dots \sum_{i_M} \sqrt{P(Y^t=y_{i_1}, \dots, Y^{t+M}=y_{i_M}|S^t)} |S^{t+M}(S^t, y_{i_1}, \dots, y_{i_M})\rangle \otimes |y_{i_1}\rangle \otimes \dots \otimes |y_{i_M}\rangle \quad (7)$$

where $|S^{t+M}(S^t, y_{i_1}, \dots, y_{i_M})\rangle$ is the quantum state that would have been prepared if the system was originally in causal state S^t then outputted string $y_{i_1} \dots y_{i_M}$, and a new causal state directly set according to this output sequence. Measuring the string of output tape subsystems thus still ensures that the internal memory state collapses into the correct causal state $|S^{t+M}\rangle$, conditional on the string observed.

In either mode of operation, it is the von Neumann entropy $-\text{Tr } \rho \log \rho$ of this first subsystem, which remains within the simulator at all times, that we consider to be the internal memory cost.

Derivation of discrete eigenspectrum. The quantum machine state corresponding to the system being in classical state α is given as $|S_\alpha\rangle = \sum_\beta \sqrt{p_{\beta\alpha}} |\beta\rangle$. Assuming $\{p_{\beta\alpha}\}_{\beta\alpha}$ is simply connected, the quantum machine will reach a stationary state $\rho = \frac{1}{N} \sum_\alpha |S_\alpha\rangle \langle S_\alpha|$. Rather than directly calculating the entropy of ρ , we can instead evaluate the entropy of the associated *Gram matrix* g , whose elements $g_{\alpha\beta}$ are given by the overlaps $\frac{1}{N} \langle S_\alpha | S_\beta \rangle$.³ As the transition probabilities are translationally invariant, it follows that the Gram matrix is *circulant* [31]. Since all rows can be derived by cyclic permutation of the top row, we shall drop one index and write the top row as $g_\alpha = g_{0\alpha}$. The eigenvalues of the Gram matrix are given by $\lambda_k = \sum_\alpha g_\alpha \exp(-\frac{2\pi i}{N} \alpha k)$ for $k = 0, \dots, N-1$, which can immediately be recognized as the discrete Fourier transform (DFT) of $\{g_\alpha\}_\alpha$, which we denote as $\mathcal{F}(g_{0\alpha})$.

Moreover, the inner product $\langle S_0 | S_j \rangle = \sum_\alpha \sqrt{p_{\alpha 0} p_{\alpha j}}$, has the form of a convolution $\sqrt{p} * \sqrt{q}$, where we have rewritten $p_{\alpha j}$ as $q_{0(\alpha-j)}$ such that q is the N -periodic extension of the reflection of p ; $q_{0j} = p_{(N-j)0}$ and $q_{0j} = q_{0(j+N)}$. We may then apply the circular convolution

theorem to find the eigenvalues of g , and therefore of ρ :

$$\lambda_k = \frac{1}{N} \mathcal{F}[\sqrt{p_{j0}}] \mathcal{F}[\sqrt{p_{(N-j)0}}]. \quad (8)$$

These eigenvalues can hence be found efficiently by numerical algorithms, such as the fast-Fourier transform.

Example: Dirac-delta shift function. Let the shift function be $P(x) = \delta(x - x_0)$ for some $x_0 \in [0, 1)$. It can be seen that all $p_{j0} = 0$ except for the one at index j' that incorporates the delta peak where $p_{j'0} = 1$. Hence, $\mathcal{F}(p_{j0}) = \exp(-2\pi i \frac{j'}{N} k)$ and $\mathcal{F}(p_{(N-j)0}) = \exp(-2\pi i \frac{(N-j')}{N} k)$, and so $\lambda_k = \frac{1}{N}$ for all k . Thus, the entropy requirement is $\log N$.

Example: Uniform shift function. Consider the uniform shift function $P(x) = 1$ for $x \in [0, 1)$. Here, $p_{j0} = \frac{1}{N}$, and so $\mathcal{F}[\sqrt{p_{j0}}] = \sqrt{N}$ for $k = 0$ and 0 for all other k . As such, we find that the eigenvalue $\lambda_0 = 1$, and all other eigenvalues $\lambda_1 = \dots \lambda_{N-1} = 0$, and hence the entropy of the Gram matrix is zero, for all values of N .

Sampling Fourier transforms. It will be useful to show an auxiliary relationship between discrete and continuous Fourier transforms. Let $g(x)$ be a function over the range $x \in [0, 1]$ that is sampled at N equally spaced points with values given by $g_n = g(\frac{n}{N})$ for $n = 0 \dots N-1$. We can construct a function $g_{\text{comb}}(x) = \sum_{n=0}^{N-1} \delta(x - \frac{n}{N}) g(x)$, whose Fourier transform is

$$\begin{aligned} \mathcal{F}(g_{\text{comb}}(x)) &= \int_{-\infty}^{\infty} dx \sum_{n=0}^{N-1} \delta\left(x - \frac{n}{N}\right) g(x) \exp(-2\pi i k x) \\ &= \sum_{n=0}^{N-1} g\left(\frac{n}{N}\right) \exp\left(-2\pi i \frac{n}{N} k\right), \end{aligned} \quad (9)$$

which when evaluated at integer k is exactly the DFT of the samples $\{g_n\}$, which we write as $\{\lambda_k\}$.

If g is periodic, it is always possible to offset the position of the sample window of g by some integer c without changing the values of g 's DFT. For the functions we consider in this article, it is more convenient to start at $-\frac{N}{2}$, since typically $g_{-\frac{N}{2}}, g_{\frac{N}{2}} \rightarrow 0$ and $g_0 = 1$. Moreover, once the sample window has been set, the values of $g(x)$ outside this window can not affect λ_k , since they do not feature in the sum. Thus, instead of considering sampling $g(x)$ across a finite window, we can consider an infinite delta train sampled at the same intervals, but across a function $g_{\text{once}}(x)$ where $g_{\text{once}}(x) = g(x)$ inside the range of the sample window (i.e. $[-\frac{1}{2}, \frac{1}{2})$ for the window used

³ This works by constructing a fictitious purification of ρ , given $|\Psi\rangle = \sum_i \frac{1}{\sqrt{N}} |S_i\rangle \otimes |i\rangle$ (where $\{|i\rangle\}_i$ is an orthonormal basis) such that $\text{Tr}_B |\Psi\rangle \langle \Psi| = \rho$ and $\text{Tr}_A |\Psi\rangle \langle \Psi| = g$. Since the von Neumann entropy of pure state $|\Psi\rangle$ is 0, it follows from triangle inequalities that $H(\rho) = H(g)$.

in this article) and $g_{\text{once}}(x) = 0$ outside this range. Here

$$\begin{aligned}\lambda_k &= \mathcal{F}(g_{\text{comb}}(x)) = \mathcal{F}\left(g(x) \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} \delta(x - \frac{n}{N})\right) \\ &= \mathcal{F}\left(g_{\text{once}}(x) \sum_{n=-\infty}^{\infty} \delta(x - \frac{n}{N})\right) \\ &= \mathcal{F}(g_{\text{once}}(x)) * \sum_{m=-\infty}^{\infty} \delta(k - mN),\end{aligned}\quad (10)$$

where we have used the convolution theorem in the final step. The periodic sampling of $g(x)$ causes the Fourier transform to be periodic with period N (a phenomenon known as aliasing), such that $\lambda_k = \lambda_{k+N}$; the convolution with a delta train effectively makes λ_k a *periodic sum* of $\mathcal{F}(g_{\text{once}}(x))$. This periodicity allows us the freedom to choose a convenient range of k . In this article, we will typically use $-\frac{N}{2}$ to $\frac{N}{2} - 1$. If $\mathcal{F}(g_{\text{once}}(x)) \approx 0$ outside the chosen range, then we can approximate

$$\lambda_k \approx [\mathcal{F}(g_{\text{once}}(x))](k). \quad (11)$$

Asymptotic limit of eigenvalues. For large N , we can derive an expression for λ_k in terms of the probability density function $P(x)$. We substitute $p_{\alpha 0}$ with $\frac{1}{N}P(\frac{\alpha}{N})$, which for Riemann-integrable $P(x)$ is an arbitrarily good approximation in the limit of $N \rightarrow \infty$. Similarly, we may substitute $p_{\alpha j}$ with $\frac{1}{N}P^\circ(-\frac{j-\alpha}{N})$, where $P^\circ(x)$ denotes the 1-periodic extension⁴ of $P(x)$. We can then see, taking the limit of the Riemann sum for a product of two functions, that $\lim_{N \rightarrow \infty} \langle S_0 | S_j \rangle = \lim_{N \rightarrow \infty} \sum_{\alpha=0}^{N-1} \sqrt{p_{\alpha 0} p_{\alpha j}} = \lim_{N \rightarrow \infty} \sum_{\alpha=0}^{N-1} \frac{1}{N} \sqrt{P(\frac{\alpha}{N}) P^\circ(-\frac{j-\alpha}{N})} = \int_0^1 dx \sqrt{P(x) P^\circ(y-x)}$, where $y = \frac{j}{N}$. Moreover, since P only has support in $[0, 1)$, we can rewrite the integral limits from $-\infty$ to ∞ , and conclude that $\lim_{N \rightarrow \infty} \langle S_0 | S_j \rangle = [\sqrt{P(x)} * \sqrt{P^\circ(-x)}](y)$ sampled at $y = 0, \frac{1}{N}, \dots, \frac{N-1}{N}$. Thus by treating g_j as samples from a function $g(y = \frac{j}{N})$ at discrete intervals of $\frac{1}{N}$, we find that $g_j \approx \frac{1}{N}g(y = \frac{j}{N})$ for large N , and hence

$$g(y = \frac{j}{N}) = \left[\sqrt{P(x)} * \sqrt{P^\circ(-x)} \right](y). \quad (12)$$

As shown in eq. (10), the eigenvalues $\{\lambda_k\}$ are given by $\lambda_k = [\mathcal{F}(g_{\text{once}}(x)) * \sum_{m=-\infty}^{\infty} \delta(k - mN)]$ evaluated at integers $k = 0, 1, \dots, N-1$, where $g_{\text{once}}(y) = g(y)$ over an (arbitrary) single period of $g(y)$ and takes the value zero elsewhere. Due to the periodic summation, it can be seen also that $\lambda_k = \lambda_{k+N}$, and so we are also free

to choose the most convenient range for k , which will typically be from $-\frac{N}{2}$ to $\frac{N}{2} - 1$. If $[\mathcal{F}(g_{\text{once}})](k) \approx 0$ when $|k| > \frac{N}{2}$, then the approximation

$$\lambda_k \approx [\mathcal{F}(g_{\text{once}})](k) \quad \text{for } k = -\frac{N}{2}, \dots, \frac{N}{2} - 1 \quad (13)$$

is reasonable. This corresponds to the case where aliased frequencies are negligible, and holds true for the examples we shall consider.

Example 1: Gaussian noise. Suppose the shift function of the particle is given by a Gaussian distribution $G_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ about μ with standard deviation $\sigma \ll 1$ such that we can ignore the probability of the particle looping around the ring.

Derivation of eigenvalues. We can express $\sqrt{G_{\mu, \sigma}(x)}$ as a Gaussian:

$$\begin{aligned}\sqrt{G_{\mu, \sigma}(x)} &= \sigma^{-\frac{1}{2}} (2\pi)^{-\frac{1}{4}} \exp\left(-\frac{(x-\mu)^2}{4\sigma^2}\right) \\ &= \sigma^{\frac{1}{2}} (2\pi)^{\frac{1}{4}} \sqrt{2}(\sqrt{2}\sigma)^{-1} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{(x-\mu)^2}{2(\sqrt{2}\sigma)^2}\right) \\ &= \sigma^{\frac{1}{2}} 2^{\frac{1}{2}} (2\pi)^{\frac{1}{4}} G_{\mu, \sqrt{2}\sigma}(x).\end{aligned}\quad (14)$$

It can be easily verified that $g_{\mu, \sigma}(-x) = g_{-\mu, \sigma}(x)$.

We also note that $\mathcal{F}(g_{\mu, \sigma}(x))$ is also Gaussian:

$$\begin{aligned}\mathcal{F}(g_{\mu, \sigma}(x)) &= \exp(2\pi i \mu k) \exp(-2(\pi\sigma)^2 k^2) \\ &= (2\pi)^{\frac{1}{2}} \frac{1}{2\pi\sigma} \exp(-2\pi i \mu k) \\ &\quad \cdot (2\pi)^{-\frac{1}{2}} \left(\frac{1}{2\pi\sigma}\right)^{-1} \exp\left(-\frac{k^2}{2\left(\frac{1}{2\pi\sigma}\right)^2}\right) \\ &= (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp(-2\pi i \mu k) g_{0, \frac{1}{2\pi\sigma}}(k)\end{aligned}\quad (15)$$

Likewise, we can express $[G_{\mu, \sigma}(x)]^2$ as a Gaussian:

$$\begin{aligned}[G_{\mu, \sigma}(x)]^2 &= \sigma^{-2} (2\pi)^{-1} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) \\ &= \sigma^{-1} (2\pi)^{-\frac{1}{2}} 2^{-\frac{1}{2}} \left(\frac{\sigma}{\sqrt{2}}\right)^{-1} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{(x-\mu)^2}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2}\right) \\ &= \sigma^{-1} (2\pi)^{-\frac{1}{2}} 2^{-\frac{1}{2}} G_{\mu, \frac{\sigma}{\sqrt{2}}}(x).\end{aligned}\quad (16)$$

Taken together (making sure to substitute in the correctly modified values of μ and σ), this allows us to provide an analytic solution for eq. (13) for Gaussian shift

⁴ Equivalent to wrapping x to $[0, 1)$ before evaluating $P(x)$.

functions:

$$\begin{aligned}
\lambda_k &= \mathcal{F}\left(\sqrt{G_{\mu,\sigma}(x)}\right) \mathcal{F}\left(\sqrt{G_{\mu,\sigma}(-x)}\right) \\
&= \mathcal{F}\left(\sqrt{g_{\mu,\sigma}(x)}\right) \mathcal{F}\left(\sqrt{G_{-\mu,\sigma}(x)}\right) \\
&= 2\sigma(2\pi)^{\frac{1}{2}} \mathcal{F}\left(G_{\mu,\sqrt{2}\sigma}(x)\right) \mathcal{F}\left(G_{-\mu,\sqrt{2}\sigma}(x)\right) \\
&= 2\sigma(2\pi)^{\frac{1}{2}} (2\pi)^{-1} (\sqrt{2}\sigma)^{-2} \exp(-2\pi i \mu k) \exp(2\pi i \mu k) \\
&\quad \cdot [G_{0,\frac{1}{2\sqrt{2}\pi\sigma}}(k)]^2 \\
&= (2\pi)^{-\frac{1}{2}} \sigma^{-1} \left(\frac{1}{2\sqrt{2}\pi\sigma}\right)^{-1} (2\pi)^{-\frac{1}{2}} 2^{-\frac{1}{2}} G_{0,\frac{1}{4\pi\sigma}}(k) \\
&= G_{0,\frac{1}{4\pi\sigma}}(k). \tag{17}
\end{aligned}$$

Hence, we see that choosing Gaussian transfer function with standard deviation $\sigma \ll 1$ corresponds to a spectrum of eigenvalues with standard deviation $\frac{1}{4\pi\sigma}$.

Upper bound on quantum memory cost. We now demonstrate that the entropy of such a system, given $H_Q = -\sum_k \lambda_k \log_2 \lambda_k$, is finite by bounding it from above. For convenience, we write $\lambda(k) := G_{0,\frac{1}{4\pi\sigma}}(k) = A \exp(-Bk^2)$ where $A = 2\sqrt{2}\pi\sigma$ and $B = 8\pi^2\sigma^2$, and will perform the calculation in units of *nats*. Thus, consider $c(k) = -\lambda(k) \ln \lambda(k)$, explicitly

$$c(k) = A \exp(-Bk^2) (Bk^2 - \ln A). \tag{18}$$

By setting $\frac{dc}{dk} = 2ABk \exp(-Bk^2) (-Bk^2 + \ln A + 1) = 0$, we find that c has stationary values at 0, $\pm\infty$ and when

$$k = \pm \sqrt{\frac{\ln A + 1}{B}} = \pm \sqrt{\frac{\ln(2\sqrt{2}\pi\sigma) + 1}{8\pi^2\sigma^2}}. \tag{19}$$

When $\sigma < \frac{1}{2e\sqrt{2}\pi} \approx 0.073$, these last two solutions disappear, and since we are in the regime of $\sigma \ll 1$, this condition is satisfied. Hence, for small σ , $c(k)$ monotonically decreases from its maximum value at $k = 0$ for both positive and negative k . This allows us to apply the Maclaurin–Cauchy integral bound (see e.g. [32]),

$$\int_m^\infty c(k) dk \leq \sum_{k=m}^\infty c(k) \leq c(m) + \int_m^\infty c(k) dk, \tag{20}$$

which holds for any monotonically decreasing region $[m, \infty)$ of a function $c(k)$ (here, $m = 0$).

Using known results for definite Gaussian integrals,

$$\int_0^\infty e^{-Bx^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{B}} \quad \text{and} \quad \int_0^\infty x^2 e^{-Bx^2} dx = \frac{1}{4} \sqrt{\frac{\pi}{B^3}}, \tag{21}$$

we evaluate

$$\begin{aligned}
\int_0^\infty c(k) dk &= AB \frac{1}{4} \sqrt{\frac{\pi}{B^3}} - A \ln A \sqrt{\frac{\pi}{B}} \\
&= \frac{A}{2} \sqrt{\frac{\pi}{B}} \left(\frac{1}{2} - \ln A \right) \\
&= \frac{1}{2} \left(\frac{1}{2} - \ln A \right). \tag{22}
\end{aligned}$$

Since $c(0) = -A \ln A$, we find from equation (19) that

$$\begin{aligned}
\sum_{k=m}^\infty c(k) &\leq \left(\frac{1}{2} - \ln A \right) - A \ln A, \\
&\leq \frac{1}{4} - \left(\frac{1}{2} + A \right) \ln A. \tag{23}
\end{aligned}$$

To obtain a bound on H_Q , we double the above since $c(k)$ is even, and multiply by $\frac{1}{\ln 2}$ to convert from *nats* to *bits* (equivalently, change the base \ln to \log_2 since $\frac{\ln x}{\ln 2} = \log_2 x$): $H_Q \leq \frac{1}{2 \ln 2} - (1 + 2A) \log_2 A$. In terms of the shift function's standard deviation σ , this gives our result

$$H_Q \leq \frac{1}{2 \ln 2} - \left(1 + 4\sqrt{2}\pi\sigma \right) \log_2 2\sqrt{2}\pi\sigma. \tag{24}$$

In the limit of small σ , the leading term of the entropy thus scales with $-\log_2 \sigma$, such that halving the width of the standard deviation adds one bit to the maximum required quantum memory cost.

Example 2: Uniform white noise. The normalized top-hat (rectangular) shift function allowing for jumps of up to $\pm\Delta$ around a constant displacement μ is written

$$S_\Delta(x) = \begin{cases} \frac{1}{2\Delta} & \text{for } \mu - \Delta \leq x \leq \mu + \Delta, \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Derivation of eigenvalues. Taking the square root of this function alters its normalization, but not its shape: $\sqrt{S_\Delta(x)} = \sqrt{2\Delta} S_\Delta(x)$.

Suppose $0 < \Delta < \frac{1}{2}$. In this case, $S_\Delta(x) * S_\Delta(-x)$ yields the triangle function

$$S_\Delta(x) * S_\Delta(-x) = \begin{cases} 1 - \frac{x}{2\Delta} & \text{for } 0 \leq x \leq 2\Delta, \\ 1 + \frac{x}{2\Delta} & \text{for } -2\Delta \leq x < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

This function is independent of the constant displacement μ . Indeed, non-zero μ only results in perfectly cancelling terms $e^{2\pi i k \mu}$ and $e^{-2\pi i k \mu}$ in the Fourier transform.

Basic Fourier analysis tells us that $S_\Delta(x)$ transforms into a *normalized sinc function* ($\text{sinc } x = \sin(\pi x)/\pi x$), and the triangle function into the square of this: $\mathcal{F}(S_\Delta(x) * S_\Delta(-x)) = 2\Delta \text{sinc}^2(2k\Delta)$. As this tends to 0 for large k , we can approximate the values of λ_k for large N using eq. (13), to find the eigenspectrum

$$\lambda_k = 2\Delta \text{sinc}^2(2k\Delta) \quad \text{for } k = -\frac{N}{2}, \dots, \frac{N}{2} - 1. \tag{27}$$

Upper bound on quantum memory cost. Through the careful deployment of mildly-intimidating algebra, we can also derive an upper bound on entropy cost of simulating the square shift function. The outline of the proof is as follows. To bound $\sum_k c(k)$ where $c(k) = -\lambda_k \ln \lambda_k$, we first construct a monotonically decreasing function $d(k)$ that satisfies $c(k) \leq d(k)$ at every k , and then show that $\sum d(k)$ is bounded from above. This sum will hence

also upper-bound $\sum_k c(k)$. As with the Gaussian example, for algebraic convenience, we will use natural logarithms and only consider the region of positive k . In the final stage, we will convert from *nats* to *bits*, and use the evenness of $c(k)$ to arrive at the full bound.

Explicitly, we write

$$c(x) = -2\Delta \frac{\sin^2 x}{x^2} \ln \left(2\Delta \frac{\sin^2 x}{x^2} \right) \quad (28)$$

where we have made the substitution $x = 2\pi k\Delta$.

In the region $x > 0$, we can expand

$$c(x) = -2\Delta \frac{\sin^2 x}{x^2} \left[\ln(\sin^2 x) - 2 \ln \left(\frac{x}{\sqrt{2\Delta}} \right) \right]. \quad (29)$$

The function $-y \ln y$ has a maximum value of $\frac{1}{e}$ at $y = e$, and so we can upper bound $c(x)$ by making the substitution of $-\sin^2 x \ln \sin^2 x$ with $\frac{1}{e}$. Since $\sin^2 x \in [0, 1]$, in the region $x > \sqrt{2\Delta}$ where $4\Delta \ln \left(\frac{x}{\sqrt{2\Delta}} \right) > 0$, we can likewise upper bound $c(x)$ by making the substitution of $\sin^2(x)$ with 1. Thus, for the region $x > \sqrt{2\Delta}$, we have a function $f(x) \geq c(x)$ given

$$f(x) = \frac{2\Delta}{x^2} \left[\frac{1}{e} + 2 \ln \frac{x}{\sqrt{2\Delta}} \right]. \quad (30)$$

However, as we plan to ultimately apply the Maclaurin–Cauchy integral convergence test, it is only convenient to use this upper bound in the region of x where $f(x)$ monotonically decreases. We identify this region by setting $\frac{df}{dx} = \frac{4\Delta}{x^3} \left(-2 \ln \frac{x}{\sqrt{2\Delta}} + 1 - \frac{1}{e} \right) = 0$, to find that $f(x)$ decreases monotonically when $x \geq \sqrt{2\Delta} \exp \left(\frac{e-1}{2e} \right)$, descending from its maximum value of $\exp \left(\frac{1-e}{e} \right)$.

However, once again consider $c(x)$. Since it has the form of $-y \ln y$, it follows that in *any* region, $c(x) \leq \frac{1}{e}$. Since $\frac{1-e}{e} > \frac{1}{e}$, we can then upper bound $c(x)$ in the region of $0 \leq x \leq \sqrt{2\Delta}$ to form the monotonically decreasing function $d(x)$ given

$$d(x) = \begin{cases} \exp \left(\frac{1-e}{e} \right) & 0 \leq x \leq \sqrt{2\Delta} \exp \left(\frac{e-1}{2e} \right) \\ \frac{2\Delta}{x^2} \left[\frac{1}{e} + 2 \ln \frac{x}{\sqrt{2\Delta}} \right] & x > \sqrt{2\Delta} \exp \left(\frac{e-1}{2e} \right), \end{cases} \quad (31)$$

that is guaranteed to satisfy $d(x) \geq c(x)$ for all $x \geq 0$. At this point, it is convenient to express this again in terms of k , making the substitution $k_{\text{split}} = \frac{1}{\pi\sqrt{2\Delta}} \exp \left(\frac{e-1}{2e} \right)$:

$$d(k) = \begin{cases} \exp \left(\frac{1-e}{e} \right) & 0 \leq k \leq \lceil k_{\text{split}} \rceil \\ \frac{1}{2\Delta\pi^2 k^2} \left[\frac{1}{e} + 2 \ln \left(\pi\sqrt{2\Delta}k \right) \right] & k > \lceil k_{\text{split}} \rceil, \end{cases} \quad (32)$$

where $\lceil k_{\text{split}} \rceil$ represents the lowest integer above (or including) k_{split} . This rounding is necessary since $k_{\text{split}} = \frac{1}{\pi\sqrt{2\Delta}} \exp \left(\frac{e-1}{2e} \right)$ is in general not an integer. To upper bound $c(k)$ at all points, we must round up this split between the regions of k , since $\exp \left(\frac{1-e}{e} \right)$ upper bounds

all $f(k)$. (I.e. being slightly too inclusive in the first region will result in a slightly higher value of $d(k)$ for the first k satisfying $k \geq k_{\text{split}}$).

Having derived our monotonically decreasing function $d(k)$, we are now in a position to show that $\sum_{k=0}^{\infty} d(k)$ is finite for $\Delta > 0$. Writing $\sum_{k=0}^{\lceil k_{\text{split}} \rceil} d(k) + \sum_{k=\lceil k_{\text{split}} \rceil}^{\infty} d(k)$ (for an upper bound, it is fine if a term is counted twice!), we evaluate the two regions separately. Firstly,

$$\begin{aligned} \sum_{k=0}^{\lceil k_{\text{split}} \rceil} d(k) &\leq \left(1 + \frac{1}{\pi\sqrt{2\Delta}} \exp \left(\frac{e-1}{2e} \right) \right) \exp \left(\frac{1-e}{e} \right) \\ &= \exp \left(\frac{1-e}{e} \right) + \frac{1}{\pi\sqrt{2\Delta}} \exp \left(\frac{1-e}{2e} \right) \end{aligned} \quad (33)$$

where we have used $\frac{1}{2\pi\Delta} \exp \left(\frac{e-1}{2e} \right) + 1 > \lceil \frac{1}{2\pi\Delta} \exp \left(\frac{e-1}{2e} \right) \rceil$. Secondly, using the Maclaurin–Cauchy integral test (see e.g. [32]), we bound

$$\begin{aligned} \sum_{k=\lceil k_{\text{split}} \rceil}^{\infty} d(k) &\leq d(\lceil k_{\text{split}} \rceil) + \int_{\lceil k_{\text{split}} \rceil}^{\infty} d(k) dk \\ &\leq \exp \left(\frac{1-e}{e} \right) + \int_{k_{\text{split}}}^{\infty} d(k) dk, \end{aligned} \quad (34)$$

where the second line follows by substituting $d(\lceil k_{\text{split}} \rceil)$ with the maximum value of $d(k)$, and by failing to round up the lower bound of the integral (thus including an extra contribution equal to $\int_{k_{\text{split}}}^{\lceil k_{\text{split}} \rceil} d(k) dk \geq 0$). This integral may be analytically solved,

$$\begin{aligned} \frac{1}{2\pi^2\Delta} \int_{k_{\text{split}}}^{\infty} \frac{1}{k^2} \left[\frac{1}{e} + 2 \ln \left(\pi\sqrt{2\Delta}k \right) \right] \\ = \left[\frac{-1}{2\pi^2\Delta k} \left(\frac{1}{e} + 2 + 2 \ln \left(\pi\sqrt{2\Delta}k \right) \right) \right]_{\frac{1}{\pi\sqrt{2\Delta}} \exp \left(\frac{e-1}{2e} \right)}^{\infty} \\ = \frac{3}{\pi\sqrt{2\Delta}} \exp \left(\frac{1-e}{2e} \right). \end{aligned} \quad (35)$$

Combining these two terms, we arrive at:

$$\sum_{k=0}^{\infty} d(k) \leq \frac{4}{\pi\sqrt{2\Delta}} \exp \left(\frac{1-e}{2e} \right) + 2 \exp \left(\frac{1-e}{e} \right). \quad (36)$$

Finally, to bound the entropy $H_Q = -\sum_{\infty} \lambda_k \log_2 \lambda_k$, we must double the above ($c(k)$ is even, and equation (36) bounds only the region $[0, \infty)$), and we convert from nats to bits (by including a factor of $\frac{1}{\ln 2}$):

$$H_Q \leq \frac{8}{\pi \ln 2 \sqrt{2\Delta}} \exp \left(\frac{1-e}{2e} \right) + \frac{4}{\ln 2} \exp \left(\frac{1-e}{e} \right). \quad (37)$$

By evaluating the constant terms, approximately,

$$H_Q \leq \frac{1.894}{\sqrt{\Delta}} + 3.067, \quad (38)$$

yielding our result.

- [1] *IEEE Standard for Floating-Point Arithmetic*. IEEE Computer Society, 2008. doi: 10.1109/IEEESTD.2008.4610935.
- [2] D. Deutsch. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 400(1818):97–117, jul 1985. ISSN 1364-5021. doi:10.1098/rspa.1985.0070. URL <http://rspa.royalsocietypublishing.org/content/400/1818/97>.
- [3] D. Deutsch and R. Jozsa. Rapid Solution of Problems by Quantum Computation. *Proc. R. Soc. Lond. A.*, 439 (1907):553–558, 1992. ISSN 09628444. URL <http://www.jstor.org/stable/52182>.
- [4] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, STOC '96, pages 212–219, New York, NY, USA, 1996. ACM. ISBN 0-89791-785-5. doi:10.1145/237814.237866. URL <http://doi.acm.org/10.1145/237814.237866>.
- [5] P. W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Comput.*, 26(5):1484–1509, oct 1997. ISSN 0097-5397. doi:10.1137/S0097539795293172. URL <http://dx.doi.org/10.1137/S0097539795293172>.
- [6] R. Cleve, A. K. Ekert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proc. R. Soc. Lond. A.*, 454(1969):339–354, Jan 1998. doi: 10.1098/rspa.1998.0164.
- [7] W. van Dam. *Nonlocality & Communication Complexity*. PhD thesis, University of Oxford, 2000.
- [8] R. M. de Wolf. *Quantum Computing and Communication Complexity*. PhD thesis, University of Amsterdam, 2001. URL <http://dare.uva.nl/record/1/194123>.
- [9] Gilles Brassard. Quantum Communication Complexity. *Foundations of Physics*, 33(11):1593–1616. ISSN 1572-9516. doi:10.1023/A:1026009100467. URL <http://link.springer.com/article/10.1023/A%3A1026009100467>.
- [10] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Physical Review Letters*, pages 105–108, 1989. ISSN 00319007. doi:10.1103/PhysRevLett.63.105. URL <http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.63.105>.
- [11] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3-4):817–879, 2001. ISSN 00224715. doi:10.1023/A:1010388907793. URL <http://link.springer.com/article/10.1023/%3A1010388907793>.
- [12] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney. Time's barbed arrow: Irreversibility, Crypticity, and stored information. *Physical Review Letters*, page 094101, 2009. ISSN 00319007. doi:10.1103/PhysRevLett.103.094101. URL <http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.103.094101>.
- [13] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8(1):17–24, dec 2011. ISSN 1745-2473. doi: 10.1038/nphys2190. URL <http://dx.doi.org/10.1038/nphys2190>.
- [14] R. Haslinger, K. L. Klinkner, and C. R. Shalizi. The computational structure of spike trains. *Neural computation*, 22(1):121–57, jan 2010. ISSN 1530-888X. doi:10.1162/neco.2009.12-07-678. URL <http://www.mitpressjournals.org/doi/abs/10.1162/neco.2009.12-07-678>.
- [15] C. R. Shalizi, K. L. Shalizi, and R. Haslinger. Quantifying self-organization with optimal predictors. *Physical review letters*, 93(11):118701, sep 2004. ISSN 0031-9007. doi:10.1103/PhysRevLett.93.118701. URL <http://www.ncbi.nlm.nih.gov/pubmed/15447385>.
- [16] J.G. Marques da Silva, J.C. Sartorelli, W.M. Gonçalves, and R.D. Pinto. A scale law in a dripping faucet. *Physics Letters A*, 226(5):269–274, feb 1997. ISSN 03759601. doi:10.1016/S0375-9601(96)00941-3. URL <http://www.sciencedirect.com/science/article/pii/S0375960196009413>.
- [17] R. W. Clarke, M. P. Freeman, and N. W. Watkins. Application of computational mechanics to the analysis of natural data: An example in geomagnetism. *Physical Review E*, 67(1):016203, jan 2003. ISSN 1063-651X. doi: 10.1103/PhysRevE.67.016203. URL <http://link.aps.org/doi/10.1103/PhysRevE.67.016203>.
- [18] Complexity analysis of the stock market. *Physica A: Statistical Mechanics and its Applications*, 379(1):179–187, jun 2007. ISSN 03784371. doi:10.1016/j.physa.2006.12.042. URL <http://www.sciencedirect.com/science/article/pii/S0378437107000271>.
- [19] C.-B. Li, H. Yang, and T. Komatsuzaki. Multiscale complex network of protein conformational fluctuations in single-molecule time series. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2):536–41, jan 2008. ISSN 1091-6490. doi: 10.1073/pnas.0707378105. URL <http://www.pnas.org/content/105/2/536.abstract>.
- [20] C. Lu and R. R. Brooks. P2P hierarchical botnet traffic detection using hidden Markov models. In *Proceedings of the 2012 Workshop on Learning from Authoritative Security Experiment Results - LASER '12*, pages 41–46, New York, New York, USA, jul 2012. ACM Press. ISBN 9781450311953. doi: 10.1145/2379616.2379622. URL <http://dl.acm.org/citation.cfm?id=2379616.2379622>.
- [21] M. Gu, K. Wiesner, E. Rieper, and V. Vedral. Quantum mechanics can reduce the complexity of classical models. *Nature Communications*, 3(762), 2012. ISSN 2041-1723. doi:10.1038/ncomms1761. URL <http://www.nature.com/ncomms/journal/v3/n3/full/ncomms1761.html>.
- [22] W. Y. Suen, J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu. The classical-quantum divergence of complexity in the Ising spin chain. page 8, nov 2015. URL <http://arxiv.org/abs/1511.05738>.
- [23] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield. Occam's Quantum Strop: Synchronizing and Compressing Classical Cryptic Processes via a Quantum Channel. *Scientific reports*, 6:20495, jan 2016. ISSN 2045-2322. doi:10.1038/srep20495. URL <http://www.nature.com/srep/2016/160215/srep20495/full/srep20495.html>.
- [24] M. S. Palsson, M. Gu, J. Ho, H. M. Wiseman, and G. J.

- Pryde. Experimental quantum processing enhancement in modelling stochastic processes. feb 2016. URL <http://arxiv.org/abs/1602.05683>.
- [25] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield. Minimized state complexity of quantum-encoded cryptic processes. *Physical Review A*, 93(5):052317, may 2016. ISSN 2469-9926. doi: 10.1103/PhysRevA.93.052317. URL <http://journals.aps.org/pr/abstract/10.1103/PhysRevA.93.052317>.
- [26] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield. The Ambiguity of Simplicity. feb 2016. URL <http://arxiv.org/abs/1602.08646>.
- [27] J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Information accessibility and cryptic processes. *Journal of Physics A: Mathematical and Theoretical*, 42(36):362002, sep 2009. ISSN 1751-8113. doi:10.1088/1751-8113/42/36/362002. URL <http://stacks.iop.org/1751-8121/42/i=36/a=362002>.
- [28] K. Wiesner, M. Gu, E. Rieper, and V. Vedral. Information-theoretic lower bound on energy cost of stochastic computation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468:4058–4066, 2012. ISSN 1364-5021. doi:10.1098/rspa.2012.0173. URL <http://rspa.royalsocietypublishing.org/content/468/2148/4058>.
- [29] S. Still, D. A. Sivak, A. J. Bell, and G. E. Crooks. Thermodynamics of Prediction. *Physical Review Letters*, 109(12):120604, sep 2012. ISSN 0031-9007. doi: 10.1103/PhysRevLett.109.120604. URL <http://link.aps.org/doi/10.1103/PhysRevLett.109.120604>.
- [30] A. J. P. Garner, J. Thompson, V. Vedral, and M. Gu. When is simpler thermodynamically better? Sep 2015. URL <http://arxiv.org/abs/1510.00010>.
- [31] Robert M. Gray. Toeplitz and Circulant Matrices: A Review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2005. ISSN 1567-2190. doi:10.1561/01000000006.
- [32] K. Knopp. *Theory and applications of infinite series*. Dover Publications, second english edition, 1990. ISBN 9780486661650.